Zero Correlation Between Old and New Residuals when Additional Observations are Incorporated into a Linear (Multiple) Regression Analysis.*

BU-210-M D. S. Robson January, 1966

ABSTRACT

If $Y_1, \cdots, Y_n$ are independent random variables with mean values

$$E\ (Y_i) = \beta_0 + \sum_{i=1}^{p-1} \beta_i\ x_{ij}$$

and common variance σ^2 then the deviations from the least squares regression function fitted to the first k observations $Y_1, \cdots, Y_n$ $(k < n)$ are linearly uncorrelated with the last n-k deviations from the least squares regression function fitted to all n observations $Y_1, \cdots, Y_k, Y_{k+1}, \cdots, Y_n$. The former deviations are also uncorrelated with the predicted values $\hat{Y}_1, \cdots, \hat{Y}_n$ of the regression function fitted to all n observations. This result has application in constructing tests of homoscedasticity.

* Biometrics Unit, Plant Breeding Department, Cornell University.

Zero Correlation Between Old and New Residuals when Additional Observations are Incorporated into a Linear (Multiple) Regression Analysis.*

BU-210-M

D. S. Robson

January, 1966

Introduction

The purpose of this note is to show that if $\hat{Y}_1, \cdots, \hat{Y}_k$ are the least squares linear predictors

$$\hat{Y}_j = \hat{\beta}_0 + \sum_{i=1}^{p-1} \hat{\beta}_i x_{ij}$$

$$= \sum_{i=1}^{k} a_{ij} Y_i$$

where

$$Y_j = \beta_0 + \sum_{i=1}^{p-1} \beta_i x_{ij} + \epsilon_j$$

and

$$E(\epsilon_j) = 0$$

$$E(\epsilon_j \epsilon_{j'}) = \begin{cases} \sigma^2 & \text{for } j = j' \\ 0 & \text{for } j \neq j' \end{cases}$$

then the residuals

$$f_1 = \left(Y_1 - \sum_{i=1}^{k} a_{i1} Y_i, \cdots, Y_k - \sum_{i=1}^{k} a_{ik} Y_i \right)$$

are uncorrelated with the least squares residuals

$$e_2 = \left(Y_{k+1} - \sum_{i=1}^{n} b_{i,k+1} Y_i, \cdots, Y_n - \sum_{i=1}^{n} b_{in} Y_i \right)$$

obtained after fitting $Y_1, \cdots, Y_k, Y_{k+1}, \cdots, Y_n$ to the same linear model. This result has application in the construction of tests of homoscedasticity.

---

* Biometrics Unit, Plant Breeding Department, Cornell University.

Uniqueness of the Least Squares Linear Predictors

The general (fixed effects) linear model is given by

$$Y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_{p-1} x_{p-1,1} + \epsilon_1$$
$$\vdots$$
$$Y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_{p-1} x_{p-1,n} + \epsilon_n$$

or

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p-1,1} \\ \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{p-1,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or, in matrix notation,

$$\underset{n\times 1}{Y} = \underset{n\times p}{X} \; \underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$

Setting the partial derivatives of $(Y - X\beta)'(Y - X\beta)$ equal to zero gives the set of "normal equations"

$$\underset{p\times n}{X'} \; \underset{n\times p}{X} \; \underset{p\times 1}{b} = \underset{p\times n}{X'} \; \underset{n\times 1}{Y}$$

for which the solution $b = \begin{bmatrix} b_0 \\ \vdots \\ b_{p-1} \end{bmatrix}$ is the vector of least squares estimators

of the regression coefficients $\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$. This general linear model includes as

special cases all multiple regression models and all "fixed-effects" analysis of variance models, though the latter are included only if we admit that the symmetric, square $(p\times p)$ matrix $X'X$ may be singular. Admitting this, we cannot then write the general solution as

$$b = (X'X)^{-1} X'Y$$

since, in general, $X'X$ will not have an ordinary inverse. However, if $X'X$ is singular (i.e., of rank < p) there exists not just one but infinitely many solutions b to the equation $(X'X)b = X'Y$, and if

$$b = \underset{p \times p}{G} X'Y$$

is one of these solutions then we shall say (after C. R. Rao) that G is a generalized inverse of $(X'X)$ provided that G also satisfies the relation

$$(X'X)\, G\, (X'X) = X'X$$

When these two relations are satisfied then we see that another solution to the equation $(X'X)b = X'Y$ is given by

$$\tilde{b} = GX'Y + \left[G(X'X) - \underset{p \times p}{I}\right] \underset{p \times 1}{z}$$

where z is any arbitrary $p \times 1$ vector. Multiplying both sides of this equation by $(X'X)$ confirms that $(X'X)\tilde{b} = X'Y$, since

$$(X'X)\tilde{b} = (X'X)GX'Y + [(X'X)\, G\, (X'X) - (X'X)]z$$

$$= (X'X)b + [0]z$$

$$= X'Y$$

Thus, for a fixed G the solution $\tilde{b}$ may be made to run through an infinity of values simply by varying the choice of z; furthermore, unlike the ordinary inverse, the generalized inverse G is not unique and may also be constructed in an infinite variety of ways. Despite this, as we shall see, the sum of squares due to regression and the sum of squares due to deviations from regression are uniquely determined by any choice of G and z; in fact, $\hat{Y} = X\tilde{b}$ is the same no matter which G and z are chosen. To show this latter fact we have

$$X\tilde{b} = XGX'Y + [XG(X'X) - X]z$$

and we now make use of the following lemma from matrix algebra:

Lemma: If $M'M = 0$ then M is null (i.e., if $M'M$ is null then M is null; similarly, if MM' is null then M is null).

Proof: Since $\text{Trace}(M'M) = \sum_{i,j} m_{ij}^2 = \text{Trace}(0) = 0$ then $m_{ij} = 0$ for all i,j.

Applying this lemma first to the coefficient of z we find

$$[XG(X'X) - X]'[XG(X'X) - X] = [X'XG'X' - X'][XGX'X - X]$$

$$= X'XG'X'XGX'X - X'XG'X'X - X'XGX'X + X'X$$

$$= X'XG'(X'XGX'X) - X'XG'X'X - (X'XGX'X) + X'X$$

and since the generalized inverse G must satisfy the relation $X'XGX'X = X'X$ we get the above expression reducing to

$$X'XG'X'X - X'XG'X'X - X'X + X'X = 0$$

implying, by the lemma, that $XGX'X - X$ is null, or

$$XGX'X = X$$

Thus, we have so far shown that

$$X\tilde{b} = XGX'Y$$

where, however, G is not unique. But we shall now show that XGX' is the same no matter what generalized inverse is used. That is, if G_1 and G_2 both satisfy our requirements for a generalized inverse then, applying the lemma again, we can show that $[XG_1X' - XG_2X']$ is null. Thus,

$$[XG_1X' - XG_2X'][XG_1X' - XG_2X']' = [XG_1X' - XG_2X'][XG_1'X' - XG_2'X']$$

$$= XG_1X'XG_1'X' - XG_1X'XG_2'X' - XG_2X'XG_1'X' + XG_2X'XG_2'X'$$

$$= XG_1(X'XG_1'X') - XG_1(X'XG_2'X') - XG_2(X'XG_1'X') + XG_2(X'XG_2'X')$$

and using the already established fact that $XGX'X = X$ for any G, we get

$$XG_1X' - XG_1X' - XG_2X' + XG_2X' = 0$$

Hence, $XG_1X' = XG_2X'$ or, in other words, XGX' is unique even though G is not. This concludes the proof that $\hat{Y} = X\tilde{b}$ is unique. In an analysis of variance situation, for example, $\hat{Y} = X\tilde{b}$ is simply the array of observed cell means.

Zero Correlation with Supplemental Residuals

We now proceed to the problem

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_{p-1} x_{p-1,1} + \epsilon_1 \\
&\vdots \\
Y_k &= \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + \cdots + \beta_{p-1} x_{p-1,k} + \epsilon_k \\
Y_{k+1} &= \beta_0 + \beta_1 x_{1,k+1} + \beta_2 x_{2,k+1} + \cdots + \beta_{p-1} x_{p-1,k+1} + \epsilon_{k+1} \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 x_{1,n} + \beta_2 x_{2,n} + \cdots + \beta_{p-1} x_{p-1,n} + \epsilon_n
\end{aligned}$$

or

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_k \\ \hline Y_{k+1} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p-1,1} \\ \vdots & & & \vdots & \\ 1 & x_{1k} & x_{2k} & \cdots & x_{p-1,k} \\ \hline 1 & x_{1,k+1} & x_{2,k+1} & \cdots & x_{p-1,k+1} \\ \vdots & & & \vdots & \\ 1 & x_{1,n} & x_{2,n} & \cdots & x_{p-1,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \\ \hline \epsilon_{k+1} \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$Y = \begin{bmatrix} Y_{(1)} \\ \hline Y_{(2)} \end{bmatrix} = \begin{bmatrix} X_{(1)} \\ \hline X_{(2)} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_{(1)} \\ \hline \epsilon_{(2)} \end{bmatrix} = X\beta + \epsilon$$

where we suppose that first β is estimated using all n observations and, next, β is estimated using only the first k observations. The problem is then to show that the last n-k deviations $Y_i - \hat{Y}_i$ obtained in the first instance are uncorrelated with the k deviations obtained in the second analysis.

We first note that for the complete set of data,

$$e = Y - \hat{Y} = Y - X\tilde{b} = Y - XGX'Y = [I - XGX']Y$$

and substituting the linear model, $Y = X\beta + \epsilon$, into this expression then

$$\begin{aligned} e &= \Big[\underset{n\times n}{I} - XGX'\Big]X\beta + \Big[\underset{n\times n}{I} - XGX'\Big]\epsilon \\ &= \Big[X - XGX'X\Big]\beta + \Big[\underset{n\times n}{I} - XGX'\Big]\epsilon \\ &= \Big[\underset{n\times n}{I} - XGX'\Big]\epsilon \end{aligned}$$

Similarly, if $\mathcal{G}$ is a generalized inverse of $X_{(1)}'X_{(1)}$ then the k residuals in the second analysis (involving only the first k observations) must take the form

$$\underset{k\times 1}{f_1} = \Big[\underset{k\times k}{I} - X_{(1)}\mathcal{G}X_{(1)}'\Big]\epsilon_{(1)}$$

In order to find the form of the last n-k residuals in the complete analysis we utilize the partitioned form

$$e = \begin{bmatrix} e_{(1)} \\ \hline e_{(2)} \end{bmatrix} = \left[\begin{pmatrix} \underset{k\times k}{I} & 0 \\ 0 & \underset{n-k\times n-k}{I} \end{pmatrix} - \begin{pmatrix} X_{(1)} \\ \hline X_{(2)} \end{pmatrix} G\left(X_{(1)}' \,\vdots\, X_{(2)}'\right)\right] \begin{bmatrix} \epsilon_{(1)} \\ \hline \epsilon_{(2)} \end{bmatrix}$$

or

$$\begin{bmatrix} e_{(1)} \\ \hline e_{(2)} \end{bmatrix} = \left[\begin{pmatrix} \underset{k\times k}{I} & 0 \\ 0 & \underset{n-k\times n-k}{I} \end{pmatrix} - \begin{pmatrix} X_{(1)}GX_{(1)}' & X_{(1)}GX_{(2)}' \\ X_{(2)}GX_{(1)}' & X_{(2)}GX_{(2)}' \end{pmatrix}\right] \begin{bmatrix} \epsilon_{(1)} \\ \hline \epsilon_{(2)} \end{bmatrix}$$

or

$$\begin{bmatrix} e_{(1)} \\ --- \\ e_{(2)} \end{bmatrix} = \begin{bmatrix} \epsilon_{(1)} - X_{(1)}GX'_{(1)}\epsilon_{(1)} - X_{(1)}GX'_{(2)}\epsilon_{(2)} \\ ---------------------- \\ \epsilon_{(2)} - X_{(2)}GX'_{(1)}\epsilon_{(1)} - X_{(2)}GX'_{(2)}\epsilon_{(2)} \end{bmatrix}$$

Thus, the last n-k residuals in the complete analysis are

$$e_{(2)} = \Big[\underset{n-k \times n-k}{I} - X_{(2)}GX'_{(2)}\Big]\epsilon_{(2)} - X_{(2)}GX'_{(1)}\epsilon_{(1)}$$

The covariance matrix of f_1 and $e_{(2)}$ is therefore

$$E(f_1e'_{(2)}) = \Big[\underset{k \times k}{I} - X_{(1)}GX'_{(1)}\Big] E(\epsilon_{(1)}\epsilon'_{(2)}) \Big[\underset{n-k \times n-k}{I} - X_{(2)}G'X'_{(2)}\Big]$$

$$- \Big[\underset{k \times k}{I} - X_{(1)}GX'_{(1)}\Big] E(\epsilon_{(1)}\epsilon'_{(1)}) X_{(1)}G'X'_{(2)}$$

and under the (homoscedasticity) assumption that the ϵ's are independent and identically distributed, $E(\epsilon_{(1)}\epsilon'_{(2)}) = 0$ and $E(\epsilon_{(1)}\epsilon'_{(1)}) = \underset{k \times k}{I}\sigma^2$ so

$$E(f_1e'_{(2)}) = - \Big[\underset{k \times k}{I} - X_{(1)}GX'_{(1)}\Big] X_{(1)}G'X'_{(2)}\sigma^2$$

$$= - \Big[X_{(1)}G'X'_{(2)} - X_{(1)}GX'_{(1)}X_{(1)}G'X'_{(2)}\Big] \sigma^2$$

and since, from our earlier results, $X_{(1)}GX'_{(1)}X_{(1)} = X_{(1)}$ then

$$E(f_1e'_{(2)}) = - \Big[X_{(1)}G'X'_{(2)} - X_{(1)}G'X'_{(2)}\Big] \sigma^2 = 0$$

and the desired result is established.

Similarly, f_1 is uncorrelated with the predictors $\hat{Y} = Xb$, since

$$X(b-\beta) = X\Big[GX'(X\beta+\epsilon) - \beta\Big]$$

$$= \Big[XGX'X - X\Big]\beta + XGX'\epsilon$$

and hence

$$\frac{1}{\sigma^2} E(f_1(b-\beta)'X') = E(f_1\epsilon'XG'X')\frac{1}{\sigma^2}$$

$$= \left[\underset{k\times k}{I} - X_{(1)}\mathscr{G}X'_{(1)}\right] E(\epsilon_{(1)}\epsilon')XG'X'\frac{1}{\sigma^2}$$

$$= \left[\underset{k\times k}{I} - X_{(1)}\mathscr{G}X'_{(1)}\right]\left[\underset{k\times k}{I} \;\vdots\; 0\right]\begin{bmatrix} X_{(1)}GX'_{(1)} & X_{(1)}GX'_{(2)} \\ X_{(2)}GX'_{(1)} & X_{(2)}GX'_{(2)} \end{bmatrix}$$

$$= \left[\underset{k\times k}{I} - X_{(1)}\mathscr{G}X'_{(1)}\right]\left[X_{(1)}GX'_{(1)} \;\vdots\; X_{(1)}GX'_{(2)}\right]$$

$$= \left[X_{(1)}GX'_{(1)} - X_{(1)}\mathscr{G}X'_{(1)}X_{(1)}GX'_{(1)} \;\vdots\; X_1GX'_{(2)} - X_{(1)}\mathscr{G}X'_{(1)}X_{(1)}GX'_{(2)}\right]$$

$$= \left[X_{(1)}GX'_{(1)} - X_{(1)}GX'_{(1)} \;\vdots\; X_1GX'_{(2)} - X_1GX'_{(2)}\right]$$

$$= \left[0 \;\vdots\; 0\right]$$

In connection with the problem of testing homoscedasticity, a useful implication of this last result is that in the normal case it still obtains if the k out of n Y_i's are selected on the basis of the predicted values $\hat{Y} = Xb$. For example, the predictors $\hat{Y}_1, \cdots, \hat{Y}_n$ may be ordered as $\hat{Y}_{[1]} \leq \cdots \leq \hat{Y}_{[n]}$, thus inducing a corresponding rearrangement of the rows in X,

$$\begin{bmatrix} \hat{Y}_{[1]} \\ \vdots \\ \hat{Y}_{[n]} \end{bmatrix} = \begin{bmatrix} X_{[1]} \\ \vdots \\ X_{[n]} \end{bmatrix} b$$

which, in turn, induces a rearrangement of $Y_1, \cdots, Y_n$ into, say, $Y_{(1)}, \cdots, Y_{(n)}$ where

$$\begin{bmatrix} Y_{(1)} \\ \vdots \\ Y_{(n)} \end{bmatrix} = \begin{bmatrix} X_{[1]} \\ \vdots \\ X_{[n]} \end{bmatrix} \beta + \begin{bmatrix} \epsilon_{(1)} \\ \vdots \\ \epsilon_{(n)} \end{bmatrix}$$

When the ϵ_i (in their original order $\epsilon_1, \cdots, \epsilon_n$) are independent and identically normally distributed then for f_1 calculated from $Y_{(1)}, \cdots, Y_{(k)}$ the preceding results concerning zero correlation (and now also independence) must still obtain.